

Makroaineisto on huono yksilötason johtopäätöksiin

SEPPO LAAKSONEN

Tatu Vanhanen on Richard Linnin kanssa julkaissut paljon älykkyyssosamäärään (ÄO) liittyen. Tässä kommentissani käytän lähteenäni erityisesti heidän uusinta kirjaansa *Intelligence – A Unifying Construct for the Social Sciences* (2012), jossa on selostettu myös melko tarkasti, miten he ovat mitanneet ÄO:n.

Kirja näyttää hyvin, kuinka paljon tekijät ovat tehneet työtä melko globaalin maakohtaisen aineiston rakentamiseksi. Se sisältää vastaavasti metodisesti melko samanlaisia sovelluksia monilta alueilta. Selittävänä tai korrelaation toisena muuttujana on niin sanottu kansallinen ÄO. Selitettäväksi muuttujaksi heille tuntuvat sopivan monet saatavissa olevat indikaattorit, kuten kansantulo (*per capita income*), poliittisten instituutioiden toimivuus ja jopa onnellisuus.¹

Vanhasen perustulos on siis, että kansallinen ÄO selittää 90 prosenttia PISA-testien tuottamista maiden välisistä eroista. Sen vuoksi koulujärjestelmän selitettäväksi jää pieni osa. Tältä voi näyttää suoraviivaisen makrotason tarkastelun perusteella, mutta minusta sellainen analyysi ei ole riittävää eikä hedelmällistäkään. Tämän pyrin osoittamaan jatkossa, jossa ydintulokset perustuvat mikrotason eli PISAn oppilastason analyysiin. Aluksi tarkastelen muutamaa kohtaa Vanhasen aineistossa, sitten esitän tiivistetysti oman analyysini tulokset ja lopuksi muiden esittämiä näkökohtia.

On syytä huomauttaa, että en ole löytänyt Lynniltä ja Vanhaselta yhtään tieteellisesti referoitua artikkelia heidän pääteemastaan.

¹ Mainitsen viimeisen siksi, että olen itsekin sitä tutkinut. Vanhasen ja Linnin esittämät maiden onnellisuutta mittaavat keskiarvot eivät ole samanlaisia kuin mitä olen estimoinut ”European Social Surveyssta”.

Älykkyyssosamäärän mittaaminen ja käyttö

Lynnin ja Vanhasen (2012) kirja ja monet muut lähteet kertovat, että älykkyyssosamäärän tuottaminen maille ei ole kiistaton asia. Lähtökohtana on eräänlainen ”Greenwich-taso” eli 100, johon kunkin maan taso kalibroidaan; tätä termiä on käyttänyt muun muassa Volkmar Weiss (2008). Osassa maista taso on siis korkeampi, osassa matalampi. Lähteet vaihtelevat. Monessa maassa on tehty otospohjaisia älykkyyden mittauksia, osassa ei. Otosten ajankohdat, laatu ja koko vaihtelevat. Tämä tarkoittaa sitä, että ÄO:n virhemarginaali (95 %:n luottamusväli) vaihtelee maasta toiseen. Tätä Vanhanen ei tuo selvästi esille. Jos ÄO-indikaattoria ei ole saatu, se on ”imputoitu” eli kopioitu samantapaiselta maalta. Imputointi voi olla perusteltua, jos luvut todella halutaan kaikille.

Kiusallisempaa on se, että lopullinen ÄO ottaa kirjan taulukon 2.1 mukaan huomioon menestymisen koulusaavutustutkimuksissa.² Muutos voi olla alaspäin tai ylöspäin. Suomen ÄO on älykkyydestin mukaan 97, mutta se nousee koulutustutkimuksissa menestymisen ansiosta 100,9:ään, jonka Vanhanen pyöristää 101:ksi. Weiss (2008) esittää vastaavia laskelmia muuta-

² Maan PISA-tuloksen muuntaminen ÄO:ksi on helppo, koska molemmat perustuvat normaali jakaumaan: PISA OECD-keskiarvolla 500 ja keskihajonnalla 100, ja ”Greenwich”-ÄO luvuin 100 ja 15. Mikrotasolla ongelman muodostavat PISA-oppilaat, jotka eivät ota koetta vakavissaan ja päätyvät alhaisiin pistemääriin. Kaikissa maissa alin luku on 50 (ÄO = 32,5) ja alin 1 % 257 (ÄO = 63,6). Wikipedian mukaan pisteet 50–69 merkitsevät lievää ja pisteet 35–49 keskitasoista kehitysvammaisuutta. PISA-tulosten yläpäässä ei vastaavaa ongelmaa ole. Maksimi on 900 (ÄO = 160), ja 99 %:n raja on 696 (ÄO = 129,4). Huomaa, että kehitysvammaiset eivät kuulu PISAtavoiteperusjoukkoon.

man vuoden takaa ja saa Suomen PISA-ÄO:ksi 103, samaan aikaan kun Lyyntin ja Vanhasen luku on 99. Hänen taulukossaan on Suomelle myös Rindermannin ÄO:n arvo 102. Näemme, että Suomen ÄO:lla on muutaman pisteen epävarmuusväli. Kaikista maista ÄO-mittauksia ei ole. Silloin luku saataneen puhtaasti koulusaavutustestistä. Tämä koskee esimerkiksi Montenegron ÄO:ta, jonka arvo on 86.

ÄO:ta on siis mahdollonta saada mitattua hyvin kaikille maille – ja sama koskee PISA-maita, vaikka niitä on vähemmän. Tämä ei ole Vanhasen tutkimuksen pahin ongelma. Näemme, että ÄO ja PISA-menestys eivät ole riippumattomia. Kun Vanhanen selittää PISA-tuloksia mittarilla, jossa on komponentteja aiemmista koulusaavutustesteistä, selittäjä on osin sama kuin selitettävä. Luulen, että Lyyntin ja Vanhasen (2012) kirjan muut maakorrelaatiot ja regressiot ovat pätevämpiä kuin tämä PISA-analyysi, koska selitettävä ja selittäjä ovat riippumattomia.

PISA-tuloksia voi olla hyväkin käyttää maan ja erityisesti mukana olevien yksilöiden osalta älykkyyden tai kyvyn mittareina. Älykkyys on merkittävä tekijä myös taloudellisen kehityksen edistäjänä, ja siksi taloustieteilijätkin ovat siitä kiinnostuneita. Merkittävän tutkimuksen tällä linjalla ovat tehneet Enrico Spolaore ja Romain Wacziarg (2009).

Suomalaisella aineistolla tutkimusta on tehnyt Roope Uusitalo (1999), joka hankki varusmiesten kykytestien (*cognitive skills*) tulokset yksilötasolla ja liitti nämä muihin mikroaineistoihin. Armeijan testit jaetaan kolmeen ryhmään, jotka osoittavat verbaalisia, aritmeettisia (numerollisia) ja loogisia kykyjä. Uusin julkaisu (Pekkala Kerr & al. 2013) osoittaa, että peruskoulu-uudistuksella on positiivinen vaikutus verbaaliseen testiin, mutta ei aritmeettiseen tai loogiseen testiin. Toisaalta peruskouluun siirtyminen paransi testituloksia silloin kun vanhemmilla ei ollut korkeasteen koulutusta. Tämä tulos osoittaa, että Suomen hyvä PISA-tulos johtuu myös hyvin toteutetusta peruskoulu-uudistuksesta. Toisaalta tutkimus kertoo, että koulutus edistää kykytesteissä menestymistä.

Nuoruusaikoina mitattua älyä tai kykyä haluttaisiin käyttää tutkimuksessa hyväksi. Minulta on esimerkiksi kysytty, miten PISA-osaamistietoja saisi tutkimuskäyttöön vaikkapa vuosilta 2000

ja 2003. Näin rekisteritietojen avulla olisi mahdollista seurata mukana olleiden vaiheita nyt jo yli 10 vuoden päähän. Aineistojen koot ovat suhteellisen suuria, ja johtopäätösten tekoon olisi edellytyksiä. En ole kyennyt vastaamaan kysyjille, vaan olen ohjannut heidät Jyväskylän yliopistoon, jossa yksilötiedot ovat. On selvää, että tutkimus vaatisi monta hallinnollista vaihetta tietosuojan takia, mutta toivon, että ne selvitetään.

Mikroaineistoilla kunnon analyysia

Vanhasen aineisto on maatasoinen, ja se sisältää 65 maata. Kyprosta en OECD:n sivulta hakemastani mikroaineistosta löytänyt, joten minulla ovat 64 maan tai talouden aineistot. Jätin pois Venäjän Permin sekä kolme Yhdysvaltojen aluetta, koska niitä ei ole Vanhasellakaan. Maatasoinen makroaineisto on huomattavasti pienempi kuin yli 450 000 oppilaan aineisto samoista maista. Hyödyllistä tietoa PISA-tulokseen vaikuttavista tekijöistä voi saada vain mikrotasolta. Makrotason yhteydethän perustuvat pienehköön aineistoon. Yleinen linja tällaisessa ”ekologisessa korrelaatioissa” on, että arvo on usein itseisarvoltaan isompi kuin mikrotasolla. Esimerkiksi ÄO:n ja matemaattisen lukutaidon korrelaatio makrotasolla on 0,93 mutta mikrotasolla 0,50. Sen sijaan matemaattisen lukutaidon ja kirjallisen tekstin ymmärtämisen välinen korrelaatio poikkeaa vähän (makro = 0,96; mikro = 0,88). Makrotasokin sopisi siten tällaiseen yksinkertaiseen analyysiin. Joissakin tapauksissa korrelaation merkikin vaihtuu. Esimerkiksi indikaattori, joka mittaa matematiikasta nauttimista, on mikroaineistossa niukasti negatiivisesti korreloitu matemaattisen lukutaidon kanssa (-0,06) mutta selkeästi positiivisesti korreloitu makroaineistossa (0,41). Molemmat ovat tilastollisesti merkitseviä, mitä ovat kaikki korrelaatiot mikroaineistossa suuren havaintomäärän takia.

En sulje pois makroaineiston käyttöä; se antaa jotain kiinnostavaa alkuanalyysiin. Mutta jos mikrotietoa on, makroaineistolla voi vain yrittää saada käsityksen riippuvuus- ja kausaalisuhteista selittävien ja selitettävän muuttujan välillä. Koekilujen jälkeen on paras siirtyä mikroaineistoon.

Tilastollisia malleja matemaattiselle lukutaidolle PISA 2012:ssa

Tulosteni tarkoitus on osoittaa, että makroanalyysi on vain pintahipaisu aiheeseen. Tilastollinen mallini on kuten Vanhasellakin lineaarinen regressio, mutta selittäjiä on huomattava määrä. Käytän kompleksisen survey-analyysin ohjelmaa eli mukana ovat oppilaan otospaino, osite ja kouluryvä. Mukana on myös koulukohtaisia muuttujia, mutta ei koulujärjestelmää koskevia. Lisäksi mallissa on oppilasta itseään ja hänen taustastaan kertovia muuttujia. Mukana on aina maa-muuttuja, joka ÄÖ:ta selittäjänä käytettäessä on Vanhasen artikkelista poimittu pistemäärä. Tämä on usein sama kahdessa tai useammassa maassa.

Vanhasen käyttämä makrotaso antaa regressiomallin selitystekeksi 87 prosenttia. Mikrotasolta laskettuna luku on paljon pienempi eli 29 prosenttia. Tulos osoittaa, että Vanhasen ”makroälykkyydellä” on teknisesti yhteyttä osaamiseen. Kiinnostavaa on verrata tätä malliin, jossa selittäjänä on maa. Käytännössä kyse on siis siitä, paljonko eri maiden keskiarvot selittävät. Selitysteke on edellistä korkeampi, noin 32 prosenttia, koska ÄÖ ja keskiarvo eivät käyttäydy yhteensopivasti. Kumpikin yhden selittäjän malli on yksilötasolla huono.

Laajan mallini kokonais selitysteke on 52 prosenttia. Tein erillisen mallin samoilla selittäjillä Suomen 8 800 oppilaalle. Koska maa selittää makromallissa paljon, jää yksittäisen maamallin selitysteke luonnollisesti alemmaksi. Suomelle se on 30 prosenttia. Malleista on syytä huomauttaa, että kaikki osaamismuuttujat ja muun muassa sukupuoli ovat täydellisesti mukana, mutta monissa selittäjissä on puuttuvaa tietoa. Tästä syystä en ottanut ihan kaikkia lupaavia jatkuvia muuttujia malliin. Toisaalta asettamalla kategorisessa muuttujassa puuttuva tieto omaksi kategoriakseen, puuttuvuutta ei teknisesti ole. Tulos on toki tulkittava oikein.

Olisi mahdollista tutkia kaikkia maita erikseen, mutta seuraavassa tiivistelmässä luettelen lyhyesti tärkeimpien selittäjien yhteydet sekä kaikkien maiden että Suomen aineistossa.

Koulumuuttujista luokan koko on hankalasti tulkittava ja vaatisi syvällisemmän tutkimisen. On nimittäin niin, että pieni luokkakoko johtaa huonompaan tulokseen, lievästi myös Suomessa. Erot ovat pieniä. Toisaalta Suomen suurin luok-

kakoko 28 on pieni verrattuna moniin muihin maihin. Siten todella ison luokan aiheuttamaa mahdollista tuloksen heikentymistä emme Suomelle voi tutkia.

Aineistossa on valmiiksi tuotettuna neljä koulun voimavaramuuttujaa, jotka koskevat opettajien moraalista tasoa, koulun fyysisiä puitteita, rahoitusta ja koulutusvoimavaroja. Näistä muut paitsi fyysiset puitteet ovat merkitseviä koko aineistolle. Suomelle merkitsevä on vain rahoitus, mutta sen vaikutus on negatiivinen. Syytä tähän pitäisi tutkia tarkemmin. Ei voi olla niin, että hyvä rahoitus on tulosta heikentävä tekijä. Muiden muuttujien ei-merkitsevyys johtuu siitä, että erot koulujen välillä ovat Suomessa pienet mutta monissa maissa suuret.

Oppilaita itseään koskevista muuttujista sukupuoli ei ole Suomessa merkitsevä, mutta kaikkien maiden aineistossa pojat ovat parempia. Sen sijaan itseluottamus selvittää matemaattisista tehtävistä on erittäin merkitsevä muuttuja molemmissa aineistoissa. Tämä siis tarkoittaa sitä, että oppilaille pitäisi löytyä rohkaisevaa kannustusta ennestä enemmän.

Mitä enemmän oppilas on poissa koulusta, sitä huonompi on osaamisen taso. Tämä on odotettava, mutta kyetäänkö siihen vastaamaan.

Jos oppilaan luokka-aste on matala, se on osoitus oppimisen hitaasta etenemisestä, koska kaikki ovat vuoden tarkkuudella samanikäisiä. Tulos on looginen eli huonompi matalammalla asteella. Tämä koskee molempia aineistoja.

Oppilaan taustan vaikutus on kaikista suurin. Enemmän koulutettujen vanhempien lapset ovat parempia kuin vähemmän koulutettujen. Tilanne on sama, jos kodin ammatillinen tausta on korkea. Maahanmuuttajien lapset ovat huonompia kuin syntyperäisten. Suomessa ero on keskimääräistä suurempi, mitä pidän huolestuttavana. Ydinperheen ja yksinhuoltajaperheen lapset ovat parempia kuin loppuryhmä eli ”uusioperheet” ja muut sekaperheet. Jos perheellä on äiti ja hän asuu kotona, tulos on parempi kuin jos äitiä ei ole tai tämä ei asu kotona.

Esikoulun käymisen hyöty näkyy selkeästi. Koko aineistossa tulos on huonoin ilman esikoulua. Suomessahan esikoulu koskee kaikkia. Kiinnostavaa on myös, että pidempi esikoulun käynti edistää osaamista.

Koemielessä oletan, että kirjoitetun tekstin ymmärtämisen testi PISAssa on yleisen älykkyyden mittari. Tällöin muuttuja voidaan lisätä se-

littäjäksi malliin, ja selitysaste nousee huomattavasti: koko aineistolle se on 87 ja Suomelle 82 prosenttia. Vastaavat regressiokertoimet (joustot) ovat 0,85 ja 0,82. Nämä kertoimet näyttävät, miten kirjoitetun tekstin ymmärtämisen testitulokset keskimäärin ennustaa osaamista matemaattisen lukutaidon testissä – sillä ehdolla, että muut mallin muuttujat ovat vakioituneet.

Johtopäätöksiä

Esittämäni suppea mikrotason analyysi vakuuttanee lukijan siitä, ettei Vanhasen käyttämällä makrotaso päästä kovin pitkälle PISA-tulosten selittämässä. Silti maiden välisiä tuloksia on kiinnostava tarkastella. Vanhasen analyysi sisältää paljon yksinkertaisuuksia siltäkin kannalta. Esimerkiksi hän ei välitä ÄÖ:n eikä PISAnkaan epävarmuuksista. ÄÖ:n virhemarginaalista en osaa sanoa, mutta PISAssa se on usein kohtuullasella (luokkaa 10–15 pistettä, Suomella ison otoskoon takia 8), koska otokset ovat standardoituja ja isoja. Mutta esimerkiksi Liechtensteinin keskiarvon 535 virhemarginaali standardin alittavan otoskoon vuoksi on niin suuri (170), että maa sisältyy samaan sijoituskategoriaan Suomen (pistemäärä = 519) kanssa. Virhemarginaalin huomioon ottaen Suomi on sijoilla 11–17 eikä sijalla 12, kuten Vanhanen toteaa.

Vanhanen on aikaisemminkin osallistunut PISA-keskusteluun. Helsingin Sanomien mielipidekirjoitukseensa (21.12.2007) hän sai professorien ryöpyn vastaansa. ”Mielenkiintoista, perusteetonta ja hyödytöntä,” Helsingin yliopiston erityispedagogiikan professori Jarkko Hautamäki kommentoi. Åbo Akademin kasvatustieteen professori Michael Uljens totesi Vanhasen olevan melko yksinäinen ajatustensa kanssa. Hän ei uskonut Vanhasen teorioiden vaikuttavan millään tavalla mihinkään. Myöskään älykkyyystutkija, Helsingin yliopiston kasvatustieteen professori Kari Uusikylä ei niellyt Vanhasen tuloksia. ”Ei ole yllätys, että älykkyysosumäärä ja matemaattisluonnontieteellinen lahjakkuus korreloivat”, hän totesi. ”Eri asia on, mitä tämä todistaa.”

Aiheeseen puuttui myös Osmo Soininvaara blogissaan (23.12.2005). Hän oli myönteisempi Vanhasta kohtaan todeten muun muassa: ”En väitä, että Vanhasen oletus vaikutuksen suunnasta olisi oikea, mutta professorien vastaväitteet ovat kyllä väärä. Tietysti voi olla, että toimittaja

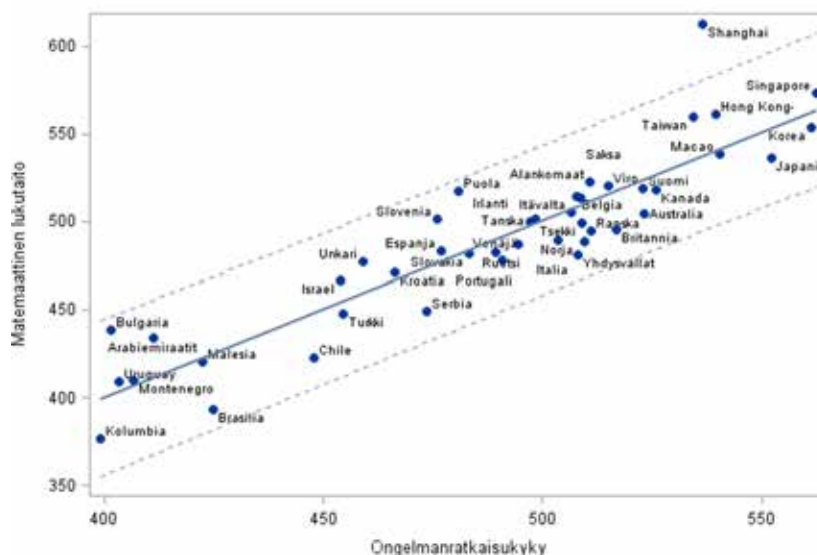
on vähän yksinkertaistanut heidän sanomansa.” Sitten hän kommentoi professorien näkemyksiä ja jatkoi yleisemmällä tasolla: ”Minua kiinnostaa kysymys, onko olemassa kansakunnan älykkyyttä asian syvemmässä merkityksessä. Kun elinkeinorakenne ja elämäntapa muuttuvat virikkeellisemmiksi ja enemmän abstraktia ajattelua vaativammiksi, seuraako tästä se, että lapsista kehittyy älykkäämpiä? Rotistakin tulee yksinkertaisessa ympäristössä tyhmiä ja virikkeisessä älykkäitä.”

Pidän näitä pohdittavina näkökohtina, mutta en jatka siitä eteenpäin. Noihin aikoihin käytiin aiheesta keskustelua myös *Tieteessä Tapahuu* -lehdessä (Töttö 2005; Ketokivi 2005). Mikko Ketokivi toi esille myös mikroaineiston tarpeellisuuden. En ole kuitenkaan havainnut, että sitä olisi sovellettu konkreettisesti esimerkiksi siihen tapaan kuin teen tässä. Toivon lisää analyysia PISAn mikroaineistosta. Jätettäköön makroaineistot vähemmälle.

Lopuksi esimerkki siitä, kuinka sironnakuvioita käytetään oikeaoppisemmin kuin Vanhanen. Olennainen tekijä on, että kuvio on tehty makrotasolle samasta mikroaineistosta eli samoista oppilaista. Mukana on kaksi osaamismuuttujaa, joista matemaattinen lukutaito on sama kuin Vanhasella, mutta uusi on huhtikuun 2014 alussa julkaistu ongelmaratkaisua mittaava muuttuja. PISAssa tähän jälkimmäiseen osallistui 43 maata, jotka kaikki ovat tässä mukana. Muuttujien välinen korrelaatio on makroanalyyseissä suurempi (= 0,91) kuin mikroanalyyseissä (= 0,84), mutta molemmat ovat isoja. Siten makro- ja mikroanalyyseissä antavat melko samanlaisen näkymän.

Kuvioon on myös piirretty yksinkertainen regressiosuora ja 95 %:n luottamusväli. Vain yksi alue (Shanghai) on luottamusvälin ulkopuolella. Siellä siis tulos ongelmanratkaisussa on suhteellisen huono. Voi olla kiinnostavaa katsoa maita, jotka ovat lähellä luottamusvälin rajaa. Se kertoo, että osaaminen näissä kahdessa osiossa ei ole kansainvälisesti katsoen hyvin balanssissa. Suomi on hyvin lähellä regressiosuoraa.

Tämä on siis makrotason tarkastelua, mikrotaso olisi kuitenkin kiinnostavampi. Huomattakoon, että kullakin keskiarvolla on luottamusvälinsä, mikä on maita vertailtaessa otettava huomioon. Näissä 43 maassa luottamusväli on melko pieni: se vaihtelee kummankin muuttujan keskiarvon osalta noin 8 pisteestä (Suomi) 34 pisteeseen



Kuvio 1. Maiden sijoittuminen ongelmanratkaisukykyyn ja matemaattisen lukutaidon keskiarvojen mukaan. Mukana myös regressiosuora ja 95 %:n luottamuvälit.

(Montenegro). Suomea on liputettu Euroopan parhaaksi ongelmanratkaisussa. Tämä on liioittelua, koska matemaattisessa osaamisessa hyvin menestynyt Sveitsi ei osallistunut

ongelmanratkaisuun. Toisaalta jos luottamusväli otetaan huomioon, Suomi on samalla tasolla Britannian, Viron ja Alankomaiden kanssa.

KIRJALLISUUS

Ketokivi, Mikko: Mittaavatko ÄO ja PISA-luvut samaa yksilön ominaisuutta? Tieteessä Tapahtuu (2005): 3, 70–71.

Lynn, Richard & Vanhanen, Tatu: Intelligence. A Unifying Construct for the Social Sciences. London: Ulster Institute for Social Research, 2012. <<http://www.ulsterinstitute.org/intelligence.html>>

Pekkala Kerr, Satu & Pekkarinen, Tuomas & Uusitalo, Roope: School Tracking and Development of Cognitive Skills. Journal of Labor Economics 31 (2013): 3.

Spolaore, Enrico & Wacziarg, Romain: The Diffusion of Development. The Quarterly Journal of Economics 124 (2009): 2, 469–529.

Töttö, Pertti: Kuvaako älykkyydosamäärä mitään? Tieteessä Tapahtuu (2005):1, 33–34.

Uusitalo, Roope: Returns to education in Finland. Labour Economics 6 (1999): 569–580.

Weiss, Volkmar: National IQ means, calibrated and transformed from educational attainment, and their underlying gene frequencies. The Mankind Quarterly 49 (2008): 2, 130–164.